1. caBIG® Life Sciences Domain Business Architecture and Domain Analysis Models

The NCI Center for Biomedical Informatics and Information Technology caBIG® program supports information integration across the bench to bedside continuum. An Enterprise Architecture Specification (EAS) is in place to promote shared development and interoperability within and beyond the caBIG® community. Key components of the architecture are business and technical models which provide a common grammar and allow traceability between business requirements, technical requirements and service specifications. EAS employs a layered approach to specification that allows both models to be constrained and extended, as appropriate, based on project-specific requirements. The Life Sciences Business Architecture Model (LS BAM) provides a common understanding to describe the business processes and operations occurring within life sciences biomedical research. The model includes business use cases and activity diagrams to represent the activities, goals, people and their interactions during the conduct of research. The Life Sciences Domain Analysis Model (LS DAM) is an object-oriented information model (class diagram) that provides a high-level common representation of the most relevant types of data generated in translational research. It is an implementation independent view of the static semantics of the domain and includes classes, attributes and associations. Life Sciences models are built with domain experts and facilitated by analysts. The LS BAM, a top down effort, is one part of an initiative to develop a complete representation of translational research encompassing basic research to clinical trials. The LS DAM effort leverages existing implementation models to create a harmonized, more abstract conceptual model. These Life Sciences models are aligned with each other and are harmonized with the Clinical Sciences domain (BRIDG DAM and the Clinical Sciences part of the caBIG® Biomedical Research BAM). Development also draws upon caBIG® standards and resources (e.g., the NCI Enterprise vocabulary services) and external standards (e.g., Health Level 7 (HL7) and Clinical Data Interchange Standards Consortium (CDISC)).

AUTHORS AND AFFILIATIONS
Lauren Becnel Boyd, PhD, Baylor College of Medicine; Lawrence Brem, SAIC; Uma Chandran, PhD, University Of Pittsburgh Medical Center; Robert Dennis, PhD, University of California at Los Angeles; Michele Ehlman, Essex Management; Elaine Freund, PhD, 3rd Millennium; Robert Freimuth, PhD, Mayo Clinic; Stephen Goldstein, Sapient Corp; Frank Hartel, PhD, NCI CBIIT; Scott Hunicke-Smith, PhD, University of Texas at Austin; Jenny Kelley, NCI LPG; Juli Klemm, PhD, NCI CBIIT; Sue Pan, PhD, SAIC; Fred Prior, PhD, Washington University in St. Louis; Konrad Rokicki, MS, SAIC; Lisa Schick, ScenPro; Mukesh Sharma, PhD, Washington University in St. Louis; Grace Stafford, PhD, The Jackson Laboratory; David Steffen, PhD, Baylor College of Medicine; Baris Suzek, MS, Georgetown University; Benjamin Tycko, MD, PhD, Columbia University

2. Test to Best - Evidence for Collaboration and Science Driven IT as Criteria for Personalized Medicine

The Ivy Genomics-Based Medicine (Ivy G.B.M.) Project is a collaborative project between nine U.S. institutions, testing whether genomic characteristics in individual brain tumors can inform optimal treatment options for individual patients. For the first time in brain cancer research, patterns of response to a wide range of targeted therapies will be analyzed against a panoply of genomic profiles assembled for each tumor. The project challenges not only many of the traditional boundaries of IT, but also the business support processes, the accelerated throughput of collaborative science, the privacy of competing pharmaceutical compounds, and the dynamism of consortia practices. Funded by the Ben & Catherine Ivy Foundation and guided by TGen, the communications and collaboration platforms needed to be quickly established, delivering support to the science, administration and operation of the serial experiments. The software allows registration, collection, tracking and aggregation of study subjects; treatments; and results; moreover, it must do so consistently across the organizations to ensure milestones are met. A combination of established platforms (Google Enterprise), open source applications (caBIG®'s caArray), and custom software (5AM's Glassbox Translational Research) were fused to meet these challenges and to support the science to be performed.

AUTHORS AND AFFILIATIONS
Michael Berens, PhD, TGen; Andrew Sloan, MD, PhD, Case Western; David James, PhD, UCSF; Howard Colman, MD, PhD; Anderson,Jann Sarkaria, MD, Mayo Clinic; Sean Lawler, PhD, Ohio State; Tom Michelson, MD, Henry Ford;

Craig Webb, PhD, Van Andel Research Institute; Yancey Gillespie, PhD, University of Alabama Birmingham; Brent Gendleman, 5AM Solutions

3. Generation of OpenClinica Electronic Case Report Forms (CRFs) using caDSR Common Data Elements via the caDSR Form Builder API

The OpenClinica Enterprise CRF Library is a Web-based repository of electronic case report forms (CRFs). It includes CRFs drawn from the NCI caDSR and uses caDSR Common Data Elements. The caDSR CRFs have been included in the CRF Library via an automated acquisition and conversion process that uses the caDSR API mapped to the OpenClinica CRF information model. By mapping OpenClinica CRF definition to CRF metadata from the National Cancer Institute's Cancer Data Standards Repository (caDSR) Form Builder, the OpenClinica Enterprise CRF Library is able to include reference metadata from source documents, such as Common Data Element identifiers, improving data harmonization and the ability to meet standardized reporting requirements.

AUTHORS AND AFFILIATIONS
Cal Collins, BA, Akasa Research; Yufang Wang, MS AKAZA Research; Krikor Krumlian, BA, AKAZA Research; Paul Galvin, BA, AKAZA Research

4. BioGrid Australia: Integrating Health and Research Data to Facilitate Improved Health Outcomes

Information technology is revolutionizing healthcare and medical research. Traditionally data was stored in a silo, making data extraction a nightmare of ethics approvals, comparability and logistics, however linking records about an individual and collating data from multiple sources has powerful potential to understand the causation of human disease and predict outcomes. BioGrid Australia has implemented technology and processes allowing researchers to extract data dynamically from multiple sources. BioGrid technology allows authorized access to securely held data to be transmitted in encrypted form and tracks and audits all data queries. It uses a federated design in which the data is stored at the owner’s institution and only extracted with authorization. Using portal and business glossary applications, researchers explore data availability and definitions before applying for access. The analytical and query tools allow researchers to extract and analyze the data themselves. Recent advances/research findings include: (1) Cartwheel.org. This international rare tumor portal allows patients to enter the details of their illness online giving them and researchers the opportunity to participate in research projects with larger groups of patients as well as facilitating enhanced information sharing and support for patients; (2) Blood parameters as biomarkers for Brian Tumours, K Field, et al.; (3) Calculating the Rapidly Escalating Cost of Treating Cancer in Australia – Time for an Increased Focus on Prevention and Screening, Ben Tran et al.; (4) Initial impact of Australia’s National Bowel Cancer Screening Program, S. Ananda et al. BioGrid is implemented in a number of states in Australia and has enabled discovery and collaborative research to be accessible via the Web while addressing security, intellectual property, and privacy.

AUTHORS AND AFFILIATIONS
Robert Merriel, BA, GD Psych, GD Acc, CPA, Melbourne Health; Marienne Hibbert, B App Sci, M App Sci, GDip Epi, GDip Stats, PhD, BioGrid Australia Ltd

5. Sharing Biorepository information through the Common Biorepository Model (CBM)

Basic life science and clinical research often uses only locally obtained specimens since there are no easy methods to search for specimens outside a lab or institution. The ability to aggregate similar specimens from various sites will expand the validation of research findings and thus, more quickly impact patient care. Here we introduced the U.S. National Cancer Institute's initiative to create a Common Biorepository Model (CBM) for biobanks and biorepository management systems vendors to use to broadcast searchable summary-level data about the available specimen collections. The goal is to reduce the time and effort required by researchers to locate a biobank that has the needed specimens. This model is part of NCIâ€™s caBIG® (cancer BioInformatics Grid)

initiatives to develop methods and models to support and fast-track research.; We have collaboratively assembled vocabulary lists from the Specimen Resource Community across additional NIH institutes that have biorepositories, as well as reached to;key biorepository management software vendors to participate in testing early CBM model, database, and associated caBIG® caGrid service to connect repositories together to publish searchable summary level deidentified information about their biobank specimen inventory. We will demonstrate the progress-to-date of the early CBM Grid service and how we are able to obtain specimen data from several test data sites.;Conclusions: The CBM initiative has drawn participation across fourteen biorepository software vendors and steps are in progress to share summary-level biorepository data across the participants using semantically integrated and syntactically interoperable methods developed for the NCI caBIG®program. Through use of the caBIG® infrastructure and the NCI Specimen Resource Locator (SRL) initiative, there is a plan to make specimen information available to researchers and institutes looking to quickly identify locations of specimens and;availability to advance their research.

AUTHORS AND AFFILIATIONS
Anna T. Fernandez, PhD, Booz Allen Hamilton, Ian Fore, DPhil, National Cancer Institute Center for Biomedical Informatics and Information Technology, Andrew Breychak, BA, Sapient Government Services, Elizabeth Prince, BS, Sapient Government Services

6. caTissue Suite: A caBIG® Open-Access, Feature-Rich Tool For Biospecimen Annotation And Data Sharing

As advances in biotechnology have increased the role of biospecimens in clinical discovery, it becomes essential that bioinformatics tools support the reporting and exchange of biospecimen information. Through the caBIG® (cancer Bioinformatics Grid) program, the US National Cancer Institute focuses on developing common vocabulary, data standards, and technology grid interfaces to enable interoperable solutions across tools required for biomedical research (clinical trials, biobanks, imaging, etc). caBIG® caTissue Suite is an open-source example that employs these standards and continually evolves to meet the biorepositories' and researchers' needs. caTissue Suite is a caBIG® biorepository management tool designed for biospecimen inventory, tracking, and annotation. The software capabilities include data migration, pathology and clinical data entry, and customization through an application programmers interface (API). caTissue is in full production (daily use) or pilot testing at several institutes inside the US and abroad, including Canada, UK, the Netherlands, and Australia. Using caGrid, multiple caTissue instances can connect to facilitate data and biospecimen sharing. A caBIG® supported, Web-based Knowledge Center (https://cabig-kc.nci.nih.gov/Biospecimen/KC) provides on-going application support via discussion forums, technical and user guides, training tools, and webinars. The growing caTissue Suite community and its requests for connecting to tools within its institutes supports the move towards a service-oriented architecture. NCI will define and use the Service Aware Interoperability Framework (SAIF) to provide secure and updated enterprise-access to clinical trial data, pathology images, and molecular analyses associated with the specimen. Other tools (including non-caBIG®) could also access the information through the service specifications and caGrid. The interoperability and re-use of data through services can save scientists research time; improve efficiency; and reduce institution and government costs.

AUTHORS AND AFFILIATIONS
Ian Fore, DPhil, National Cancer Institute Center for Biomedical Informatics and Information Technology; Anna T. Fernandez, MS, PhD, Booz Allen Hamilton, caBIG® Tissue/Biospecimen and Technology Tools Knowledge Center, Washington University

7. Comparative Analysis of Tissue Microarrays using Grid and High-Performance Computing Technologies

In this poster we present a Grid-enabled framework that is designed to support rapid, comparative analysis of expression patterns in cancer tissue microarrays (TMAs). Our framework implements a suite of image processing operations, including automated registration, segmentation, feature extraction, and classification, developed as a

Java library and C/C++ programs. Computationally intensive methods and workflows are converted to a caGrid analytical service to support collaborative projects and large-scale analyses involving hundreds or thousands of TMA disc images. The analytical service leverages high-performance computing systems, while providing service interfaces to remote clients. Images submitted to the analytical service are processed on a parallel machine and the results are returned to the client through Grid service interfaces. We also have developed a data model, called PAIS, to capture analysis results for both TMA and whole-slide microscopy images. This model captures metadata about whole-slide and TMA images and image collections, image regions and annotations from human or computerized analysis of images, and derivation history of a markup or annotation, such as algorithm information, parameters, and inputs. We have developed a caGrid-data service to host data conforming to this model. This data service enables sharing of analysis results among collaborating institutions and facilitates queries for exploration of results and comparison of results from two algorithms. The poster describes the overall framework, the various data and analytical services implemented in the framework, and present how these services are used for analysis of tissue microarrays.

AUTHORS AND AFFILIATIONS

Center for Comprehensive Informatics, Emory University

David J. Foran, PhD, The Cancer Institute of New Jersey, UMDNJ-Robert Wood Johnson Medical School,Lin Yang, PhD, The Cancer Institute of New Jersey, UMDNJ-Robert Wood Johnson Medical School,Wenjin Chen, PhD, The Cancer Institute of New Jersey, UMDNJ-Rober

8. The Cancer Therapy Evaluation Program-Enterprise System (CTEP-ESYS) Service Oriented Architecture

The Cancer Therapy Evaluation Program (CTEP) Enterprise System (CTEP-ESYS), consisting of 22 applications developed over a 12 year period using different technologies, is designed to enhance the scientific and administrative aspects of cancer clinical trial development processes. The implementation of the Service Oriented Architecture (SOA) provides continuous improvement of these processes by enhanced interoperability and data sharing across heterogeneous systems and supports collaboration with the research community through convergence with the Cancer Biomedical Informatics Grid (caBIGÂ®).;The CTEP Enterprise Services (CES) is a logical grouping of services developed using SOA to achieve interoperability in the clinical trials space. The CES uses remote services and messaging interfaces to harmonize person and organization information between CTEP and caBIGÂ® in a HL7/BRIGD compliant format and also utilizes the messaging API and caGRID services provided by caBIGÂ® to aid in the reverse synchronization of organization information from caBIGÂ® to CTEP. ;The Online Agent Order Processing (OAOP), the first application developed using SOA framework, provides workflow management and integration with shipping services provided by FEDEX for order shipment processes. The application eliminates errors during drug ordering and shipping and enhances communication between drug requesters and CTEPÂ's Pharmaceutical Management Branch (PMB) through automatic notifications. ;The Adverse Events Expedited Reporting System (AdEERS) web service, exposed by CES, enables users to submit adverse event reports from any Electronic Data Capturing (EDC) system and is currently being used by the Cancer Adverse Event Reporting System (CaAERS) to submit adverse event reports to CTEP. ;Through adoption of SOA technologies and convergence with the caBIGÂ®, the CTEP-ESYS can deliver additional benefit, such as connecting scientists and practitioners through a shareable and interoperable infrastructure and facilitating implementation of standard rules and common languages as defined by caBIGÂ® for data sharing.

AUTHORS AND AFFILIATIONS

Steven Friedman, MHSA, NCI, DCTD, CTEP, George Redmond, CGEIT, MSc, MBA, NCI, DCTD, CTEP, Shanda Finnigan, RN, BSN, CCRC, NCI, DCTD, CTEP, Yogi Byreddy, BS, PMP, CTIS, Inc., Hariprasad N Davanagere, MS, CTIS, Inc., Ganesh Panathula, BE, CTIS Inc.

9. Safety Profiler

Safety ProfilerTM, developed by CTIS Inc., is a wireless reporting solution built upon Windows Mobile 6.0 platform, has strong implications for improving the efficiency and completeness of toxicity data collection for clinical trials conducted at various levels including NIH/NCI, pharmaceutical industry, and locally funded studies. The features included in Safety ProfilerTM assist in identifying adverse event (AE), including toxicity assessment, AE grading, and reporting and maintains regulatory compliance the through use of M1/MedDRA, CTCAE, 21 CFR Part 11, ICH-GCP and HIPAA standards. Two primary systems features, data validation at the point of care and data standardization, are backed by AE term data libraries and CTC v2.0 and CTCAE v3.0/v4.0 dictionaries and incorporated into the tool to assist users while entering data in standard language. User capabilities are further enhanced through the availability of cross-mapped full terms, short names, and medical codes that are part of CTCAE v3.0/v4.0. Validation of the data as the user generates the AE information ensures that the data is relevant, compliant, and that the report is complete leading to faster, more accurate reporting, and to more timely NCI or FDA action in response to SAE reports. The Safety ProfilerTM product also bears MicrosoftÂ's approval as a Â"Designed for Windows MobileÂ" application.

AUTHORS AND AFFILIATIONS
Rajiv Shahi, MSc, PMP, MCAD, CTIS Inc.

10. imHealthe

The imHealtheTM application, developed by CTIS Inc., is an interoperable Personal Health Record (PHR) software solution currently available on both iPhone OS and Android OS platforms and is designed to overcome the many bottlenecks and redundancies existing in todayÂ's health care system that limit real-time access to individual patient health information. imHealtheTM digitally captures patient health information and allows real-time exchange between different web-based PHRs and health care providers and helps eliminate the need for redundant patient data entries and information assembling in single repositories resulting in improved efficiency and quality of patient care with reduced cost and errors. imHealtheTM combines the efficiency of a PHR (i.e., appointment scheduling, care management, reporting, billing, etc.) with the accessibility of computer-based records from multiple locations via a secure network. By using imHealtheTM, providers can, regardless of where they are, readily access medical histories, lab test results, digitized diagnostic imagery and reference databases and overcome obstacles associated with globalization and adoption of PHRs in the health care domain. imHealtheTM is designed to seamlessly integrate with commercial PHR repositories, such as Google Health, Microsoft Health Vault, and Dossia and can interact with clinical research related repositories such as NCIÂ's Patient Outcomes Services or other Hospital repositories. imHealtheTM supports patients by accurately documenting and communicating patient data and history to doctors and other healthcare providers via the convenience of mobile device and greatly reduces the potential for error or unrecorded data. imHealtheTM is easily adopted and adds value as a monitoring and tracking tool to provide the right information to the right person, anytime, in the palm of their hands.

AUTHORS AND AFFILIATIONS
Rajiv Shahi, MSc, PMP, MCAD, CTIS Inc.,Vivek Pillai, MS, MBA, PMP, CTIS Inc.,

11. Implementing Standards for Pediatric Pharmacology Research Unit Clinical Data Collection and Sharing

KAI Research, Inc. (KAI), as the Pediatric Pharmacology Research Units (PPRU) Coordinating Center, has worked with the PPRU investigators to develop a useful research resource for the Institute and its investigators: the PPRU Toolbox and Pediatric Clinical Data Repository (PeDaR). Together they support the study life cycle from protocol development through analysis.;;The PPRU Toolbox is a central resource of templates and guidelines, including templates for protocol development, standardized data elements and data collection forms, manuals of procedures (MOP), as well as tools for study and data management. The PPRU convened working groups of

experts to participate in the identification of common data elements (CDEs). The PPRU Toolbox currently contains 242 CDEs and 25 common forms with instructions for pediatric clinical studies. Three recently funded PPRU clinical studies were successfully implemented within 4 weeks using the Toolbox. Thus, start-time was reduced and at the same time data quality was increased. Based on the study protocols, common forms were customized by adding either protocol specific elements or removing optional data elements. In addition, the PPRU Toolbox helps educate new clinical investigators and study coordinators in the study development process. ;PeDaR provides a secure storage and retrieval environment for PPRU investigator initiated study data using a common set of terms and definitions to promote data sharing and metadata analysis. It contains data and protocol descriptors to facilitate queries and analyses across studies. In addition, a user friendly repository Web portal was developed and implemented. KAI has worked with the PPRU community to gather and develop requirements on the user interface for Query and Analysis tools. With over 10 PPRU legacy studies mapped to PeDaR, we developed multiple dynamic queries and graphical reports related to subject profiles, pharmacokinetics and safety evaluation.;This presentation will describe the processes for implementing standards to facilitate pediatric studies.

AUTHORS AND AFFILIATIONS
Patti Shugarts, B.S., KAI Research, Inc., Ben Piper, B.S., KAI Research, Inc., Yun Lu, Ph.D., KAI Research, Inc., Selma Kunitz, Ph.D., KAI Research, Inc., Rene Kozloff, Ph.D., KAI Research Inc.

12. Interoperable Biomedical Informatics Infrastructure

Integrative genomics provides unprecedented power to increase our understanding of basic biological processes and to determine the mechanisms of disease. This approach combines evidence from multiple data modalities such as gene expression, copy number, epigenetic, and mutation data to find the genomic causes of a disease state. It has resulted in the identification of novel mutations, the discovery of causal relationships between genomic aberrations and clinical pathologies; it has produced other important insights in the short time it has been in practice. A key catalyst in this field has been next-generation sequencing (NGS) technologies, which make possible a wide variety of genomic investigations not previously imaginable, such as the sequencing of great volumes of genomic data at rapid speed and relatively low cost. To take advantage of this wealth of data, new tools are needed that can span data modalities and support the very large datasets characteristic of integrative efforts. The Broad Institute has produced a number of software tools to facilitate integrative genomics investigations, including GenePattern, a suite of over 120 tools for the analysis of gene expression, copy number, proteomics, flow cytometry, and other data, along with extensive capabilities for combining these tools to create complex, reproducible methodologies; the Integrative Genomics Viewer (IGV), a flexible, scalable, high-performance tool for the concurrent visualization of multiple large scale datasets; and the Integrative Genomics Portal, a customizable portal for sharing, browsing, and analyzing annotated, multi-modal genomics datasets.

AUTHORS AND AFFILIATIONS
Jill P. Mesirov, PhD, The Broad Institute; Michael Reich, PhD, The Broad Institute; Ted Liefeld, MS, The Broad Institute; Jared Nedzel, MS, The Broad Institute; Jim Robinson, PhD, The Broad Institute

13. Leveraging cancer Bioinformatics Infrastructure Objects (caBIO) for research and experimental annotations

Conducting biomedical research requires access to experimental data as well as associated molecular annotations. Annotations providing detailed information on the molecular origin, biological process, and genetic alterations of associated experiment data can provide important insight on experimental outcomes. Annotating experiments with biological information requires access to data from a variety of disparate data sources in an integrative view.;;cancer Bioinformatics Infrastructure Objects (caBIO) provides Application Programming Interface (API) access to an integrative view of molecular annotations originating from a variety of data sources. Information from key annotation providers Â– including UniGene, Entrez Gene, and UniProt - is stored in the caBIO infrastructure and is updated on a semi-monthly basis. Recent additions to caBIO have included annotations from

popular microarray platforms, curated information from the Cancer Gene Index (CGI), and pathway interactions from the Pathway Interaction Database (PID). ;;caBIO data is made accessible through a variety of APIs including the caBIG grid (caGrid) API and the RESTful, SOAP, and Java APIs. caBIO is also searchable through a variety of graphical user interfaces including: a caBIO Home Page which provides a utility (FreestyleLM) for performing Â"GoogleÂ®-likeÂ" searches; a caBIO Portlet for easily accessing caBIO data via the caGrid Portal by performing pre-defined templated searches; and a caBIO iPhone App for retrieving molecular annotations via the iPhone, iPod, and iPad devices.;;caBIGÂ® translational research, knowledge discovery, and genomic analysis tools like caIntegrator, Rembrandt, geWorkbench, and other biomedical applications leverage caBIO for genomic annotations. Use of the caBIO API avoids duplication of annotation data and reduces the need for additional data management resources supporting annotations. The caBIO project is currently engaged in a pilot effort involving the development of a molecular annotation service leveraging the NCI Enterprise Conformance and Compliance Framework (ECCF).

AUTHORS AND AFFILIATIONS
Juli Klemm, PhD, NCI CBIIT, Avinash Shanbhag, PhD, NCI CBIIT, Jim Sun, SAIC, Konrad Rokicki, SAIC, Sharon Gaheen, MBA, SAIC, Liqun Qi, SAIC

14. Experiences in selecting and installing the caBIG, bioinformatics system at a new research unit in a developing country

Introduction: Science and technological advances have created unfathomable opportunities for managing and improving health in ways that were not possible previously, yet their potential for the benefit of society has not been fully realized. Important data and information are often 'trapped' in silos among research communities that are not aware of new bodies of knowledge across disciplines, do not have access to data and/or have not fully embraced the cooperative spirit of sharing. One of the objectives of The Aga Khan University's new Clinical Research Unit (CRU) is to provide a harmonized and interoperable bioinformatics environment that will empower its research community to manage, share, examine and apply existing and new knowledge in meaningful and innovative ways across disciplines at its international
campuses.
Objective: To share the experiences, processes, and resources required in selecting and installing an informatics system at a Clinical Research Unit in a developing country
Method: The process comprised of the following:
System research and selection: establishing a multidisciplinary team, defining assessment parameters, documenting work and data flow, researching and evaluating software systems
Planning and preparation: Team planning and prioritization, studying system documentation, establishing technical support link with caBIG team, progress review
System installation: Hardware configuration; sequential installation and integration of C3PR, CAAERS, PSC, Labviewer and Open Clinica on a Windows platform. Each step included trouble shooting and local testing. Timely technical guidance of the caBIG team and the Knowledge Center were instrumental in resolving challenges.
Results: caBIG was selected on the basis of several key parameters such as functionality, flexibility, cost, standardization, continuing development, technical support and the caBIG community. System modules were installed and integrated successfully on a Windows platform over four months. A total of 892 hours were spent in this first phase constituting 21% for research and evaluation of systems, 13 % for planning and progress review and 66% for system installation, problem solving and testing.
Discussion: Successful installation in large part was attributed to the perseverance of a dedicated team and timely access to technical support. Institutions should be prepared to ride the storm of unexpected
challenges during system set-up by committing adequate resources, especially dedicated staff time. The experience of the CRU has been positive, positioning it to commence the next phase of implementation involving roll out to its users.

AUTHORS AND AFFILIATIONS

Rozmin Jamal, Aga Khan University; Mohsin-e Azam, Aga Khan University; Naeem Charolia, Aga Khan University; Sajida Parveen, Aga Khan University

15. Integrative Analysis of Image and Molecular Data for Study of Brain Tumors

Joel Saltz, PhD, MD, Emory University; Daniel Brat, MD, PhD, Emory University; Jun Kong, PhD, Emory University;

We will present the methods and tools employed in study of brain tumors using integrative in silico experiments using publicly available datasets. These experiments are carried out by the In Silico Brain Tumor Research Center (ISBTRC) based at the Emory Center for Comprehensive Informatics in collaboration with Henry Ford Hospital, Stanford University and Thomas Jefferson University. The ISBTRC is funded by the NCI caBIG® In Silico Research Centers of Excellence program and dedicated to exploring novel ideas in hypothesis-driven, integrative translational research on brain tumors. The ISBRTC in silico experiments make use of complementary microscopy imaging, radiology imaging, molecular, and clinical datasets. In this project, we are developing workflows and methods for computerized analysis of microscopy images and molecular data. These methods and workflows are also being implemented as services using caGrid-enabled imaging middleware components. Results from microscopy image analyses are image markups, segmented regions (e.g., nuclei, regions of blood vessels), and annotations on these regions. Computer analysis results as well as markups and annotations done by humans are stored in a caGrid-data service using a microscopy image analysis model, called PAIS. This service enables queries for exploration of analysis results, comparisons between results from different algorithms and human markups, integration with molecular analysis results, and sharing of results with other groups. Similarly, radiology image markups and models are modeled using the caBIG® Annotation and Image Markup (AIM) model and will be stored in AIME, a caGrid-data service hosting data conforming to the AIM model. In this poster, we will describe these methods and tools and present how they are used to support integrative in silico studies and facilitate sharing of data and analytical resources.

AUTHORS AND AFFFILIATIONS

Lee Cooper, PhD, Emory University; David Gutman, MD, PhD, Emory University; Fusheng Wang, PhD, Emory University; Sharath Cholleti, PhD, Emory University; Carlos Moreno, PhD, Emory University; Tom Mikkelsen, MD, Henry Ford Hospital; Chad Holder, MD, Emory University; Patrick Widener, PhD, Emory University; Adam Flanders, MD, Thomas Jefferson University; Erwin van Meir, PhD, Emory University; Daniel Rubin, MD, MS, Stanford University; Tony Pan, MS, Emory University; Ashish Sharma, PhD, Emory University; Tahsin Kurc, PhD, Emory University

16. The Integrated Platform for Agents and Diseases (IPAD)

For over a decade, the Cancer Therapy Evaluation Program (CTEP) Enterprise System (CTEP-ESYS), which is designed to enhance the scientific and administrative aspects of cancer clinical trial development and management, has collected and maintained clinical protocol and patient data. The Integrated Platform for Agents and Diseases (IPAD) component of the CTEP-ESYS is an intuitive tool to improve data search capabilities in order to assist NCI personnel in data analysis, data mining and transactional computing. IPAD provides flexible query, analytical and integrated data warehouse capabilities and is comprised of an enterprise search engine built on a suite of components designed for aggregating information across various data sources such as the CTEP-ESYS, Enterprise Vocabulary Services (EVS), other DCTD programs (e.g., DTP, CIP, etc.), PubMed, and ClinicalTrials.gov, etc. It is a unified query tool that allows NCI/CTEP, pharmaceutical companies, Cooperative Groups, cancer centers, consortia and academia involved in developing new and novel cancer therapies to search structured or unstructured data stored in file form from the network, biomedical articles, and life science journal abstracts from the PubMed repository and other registries. IPAD users have the ability to drill down the results returned from queries to granular detail and to provide reports for tracking the timeliness of protocols in support of the Operational Efficiency Working Group (OEWG) recommendations. Information is presented in formats such as reports, graphs, charts, etc. IPAD's customizable user interface enables users to set preferences and generate reports and outputs to aid in analysis. Reports/graphs can be exported to different applications such as Word, PowerPoint, etc. IPAD's architecture is scalable and can be easily integrated with existing/future applications and

data repositories. Decision-making abilities will improve through features and capabilities that enable aggregate analysis, profiling, trending, and streamlined management of researcher's clinical trials portfolio.

AUTHORS AND AFFILIATIONS
Steven Friedman, MHSA, NCI, DCTD, CTEP; George Redmond, CGEIT, MSc, MBA, NCI, DCTD, CTEP; Shanda Finnigan, RN, BSN, CCRC, NCI, DCTD, CTEP; Sudhir Raju, MBA, MS, PMP, CTIS, Inc.; Amit Vengurlekar, MS, PMP, CTIS Inc.; Sanjeev Kakar, BS, CTIS, Inc.

17. Protein-ligand Interactions Evaluation Using Distributed Volunteer Computing Project Drugdiscovery@Home

It is well known that the drug development process from conception to commercialization has consumed increasing amounts of time and money over the past several years. Computer-aided drug discovery methods represent a promising way to reduce the cost and time of drug discovery. Studies of molecular mechanisms of drug-target interactions are hopeful not only for selective hit identification, but also at stages of hit-to-lead optimization and of predicting side-effects by reverse screening methods which can be important even for modeling of clinical trials results. The main problem with protein-ligands interactions' evaluation is the trade-off between accuracy of methods and computational costs of their usage. In our project, we have made an effort to benefit from continually increasing power of personal computers as well as from the increase of number of PC users with high-speed internet connection worldwide. This has enabled us to implement several rather computationally expensive methods such as molecular dynamics or flexible docking for high throughput screening of chemical libraries. All software implemented in DrugDiscovery@Home is open source and available for academic and industrial researcher purposes. The Autodock software was used both for the virtual screening through one protein targets conformation as well as for proteins ensemble docking. The GROMACS software was used for molecular dynamics. For the protein-ligand parametrization in GROMACS, we have used generalized amber force field (GAFF). An open source script, called ACPYPE, was written for the conversion of amber topology and coordinate files for small molecules and protein-ligand complexes. The resulting workflow was tested for evaluating protein-ligand interactions of several biotargets, which participate in Wnt signalling pathway and have ligands with known activity. We discuss details regarding how we resolved technical issues concerning project setup, software implementation, protein-ligand complexes parametrization in GROMACS and protein-ligands interactions.

AUTHORS AND AFFILIATIONS
John Shultz, Digital BioPharm Ltd.; Alan Wilter, University of Cambridge, Department of Biochemistry; Andrey Voronkov, Moscow State University, Department of Chemistry, Russian Federation, Digital BioPharm Ltd.

18. caBIG® Integration Hub

caBIG® Integration Hub (iHub) is a service oriented, robust, configurable messaging infrastructure for exchanging information among various applications and systems. It provides support for reliable messaging, reliable transactions, transformations, email notifications, payload validation, audit logging, and high availability. iHub provides the infrastructure for integrating between grid, non-grid and multi-grid environments; it is the integration engine for caBIG® Clinical Trials Suite (Suite). The Suite leverages iHub for integration among the following suite applications: caBIG® Central Clinical Participant Registry (C3PR), caBIG® Adverse Event Reporting System (caAERS), LabViewer, Patient Study Calendar (PSC), and Cancer Central Clinical Database (C3D). In addition, the Suite uses the iHub for integration with NCI enterprise services under Clinical Trials Reporting Program (CTRP). As more enterprise services become available in future, iHub will facilitate integration with those services as well. The Suite also uses iHub for uniform integration with Clinical Data Management Systems (CDMS) via the Clinical Connector. iHub is based on open source technologies such as Apache Servicemix and caBIG® infrastructure tools (e.g., Common Security Module (CSM), Common Logging Module (CLM), and Log Locator Tool (LLT)). iHub leverages caGrid security infrastructure services such as Credential Delegation Service (CDS), Dorian and Authentication services to provide seamless integration and traceability across disparate applications and scenarios.

AUTHORS AND AFFILIATIONS
Santosh Joshi, Ekagra Software Technologies; Harsh Marwaha, Ekagra Software Technologies; Ajay Nalamala, Ekagra Software Technologies

19. xService

We present the xService infrastructure that provides support for building caGrid data services from XML schemas and XML database backends. Generally a Data Service shares to the grid data that is stored in a relational database backend. caGrid data services for relational database backends are supported by caCORE SDK and caGrid caCORE service plug-ins. The xService infrastructure has been developed in response to the requirements associated with managing, exposing, and sharing datasets that conform to existing XML schemas like HL7AECG and PAIS and was driven mainly by use cases in the CardioVascular Research Grid. XML-base caGrid Data Services can be created using the xService extension for the Introduce toolkit. However, to ensure full caBIG® compatibility of these services and their data models, object-oriented models served by such services should be semantically annotated and registered in the environment. In this poster, we will present the core xService infrastructure and describe the tools we have developed to; (1) create UML models from XML schemas; (2) convert these UML models to domain models used by the Introduce toolkit for service creation; and (3) map CQL queries against the UML models to XPath queries to backend XML schemas and databases. By creating UML models from XML schemas, we are able to leverage the caBIG® model curation and annotation tools such as SIW. Our UML generation workflow adds special tags to a UML model (more specifically to the corresponding XMI document) so that the mapping between the UML model and the XML schema, from which the UML model is generated, is maintained. This is necessary as the UML model may be modified during the caBIG® semantic annotation and harmonization process. These special tags are also used when generating domain models for caGrid data services and for correctly mapping CQL queries to backend XPath queries.

AUTHORS AND AFFILIATIONS
Tony Pan, MS, Emory University; Fusheng Wang, Emory University; Tahsin Kurc, Emory University; Cristobal Vergara Niedermayr, Emory University; Ashish Sharma, PHD, Emory University; Joel Saltz, MD, PhD. Emory University

20. Automated RESTful Interface Generation for caGrid Services

caBIG® leverages SOAP-based grid services in its Service Oriented Architecture to support interoperability between system components. While SOAP service APIs provide a high degree of extensibility and syntactic conformance, their use require significant development effort and is platform dependent. The RESTful Web service paradigm represents an alternative approach to sharing resources that leverages common Web technologies like http. RESTful APIs are easier to incorporate as they use http operations, and they can be invoked directly from Web browsers. The objective of the RESTful Source code Generator Tool (RSGT) project is to allow automated generation of RESTful frontend for an existing SOAP based caGrid Service in order to provide an alternative access method to the grid-enabled data. The caGrid Introduce toolkit allows developers to generate skeletons for SOAP-based grid services by specifying the service data types and operations. The developers then provide implementation for and deploy these services. The RSGT uses the service specification and the data model to generate a RESTful Web service implementation. RSGT automatically maps RESTful API operations to each of the grid service operations. RSGT also creates additional RESTful API urls for accessing the classes defined in the domain model. The generated RESTful service translates http operation into caGrid service operation invocations and presents the results in the user requested formats, including customizable formats through XSLT. Features include: Automatic source code generation and deployment of RESTful services; support for query results formats (XML, HTML, CSV, and JSON); support for native CQL queries for advanced users; eclipse plug-in for graphical RSGT configuration; and RESTful service generation.

AUTHORS AND AFFILIATIONS

Nadir, Saghar, BS, Georgia Institute of Technology; Tony, Pan, MS, Emory University Center for Comprehensive Informatics; Ashish, Sharma, PhD, Emory University Center for Comprehensive Informatics; Joel, Saltz, MD, PhD, Emory University Center for Comprehensive Informatics

21. Using a Co-Occurrence Network of Life Science Concepts (NLSC) to search PubMed

As of 2009, approximately 90% of the over 18M articles included in PubMed had been annotated using the Medical Subject Headings controlled vocabulary (MeSH). If we consider any 2 MeSH concepts to be ''connected'' when both are chosen to annotate the same article, we can build a co-occurrence network by aggregating individual connection counts and applying various algorithms to assign a ''strength'' or ''weight'' to each connection. Such a network can be ''explored'' using a Web-based interface that displays all, or a subset of, possibly rank-ordered concepts connected to any user-specified concept, and offers links that launch PubMed searches into NCBI's Entrez interface. A prototype network and interface are currently operational. Reviewers who wish to try the current interface may use this temporary URL: http://discern.uits.iu.edu:8421/NLSC.html. Note that this network is independent of, or orthogonal to, the MeSH ontology hierarchies. In fact, this same approach could be used to find relationships among articles ''tagged'' informally by readers, or by considering words in raw text as concepts. In general this poster illustrates a method for automatically discovering relationships among concepts. This method is a novel interface for searching annotated information repositories, irrespective of the annotation vocabulary being used.

AUTHORS AND AFFILIATIONS
Michael Grobe, MS, Pervasive Technology Institute, Indiana University and Indiana Clinical and Translational Sciences Institute of the Indiana University School of Medicine

22. Harmonization of the National Institute of Neurological Disorders and Stroke (NINDS) Core Common Data Element (CDE) Case Report Form (CRF) Modules into the National Cancer Institute (NCI) cancer Biomedical Informatics Grid (caBIG®), cancer Data Standards Registry and Repository (caDSR)

In an effort to reduce study start-up time and accelerate data sharing in neurology, the National Institute of Neurological Disorder and Stroke (NINDS) started the CDE Project four years ago. This effort has resulted in the development of “General” core CDEs, which are commonly collected in all clinical studies regardless of therapeutic area, and more recently, the development of disease-specific CDEs. Meanwhile, NCI invested in a system of standards and a data warehouse structure that would allow access to and contributions from researchers sponsored by the NIH Institutes and other national and international institutions. The importance of data integration is critical because data collected at different agencies, institutions, and clinics cannot be compared or analyzed in aggregate. With the goals of data harmonization and integration in mind, the NINDS CDE Team (NCT) worked with the caBIG® data standards team to perform a gap analysis of the structure and definitions of the NINDS CDEs compared to the NCI caBIG® standards. Tools from the caBIG®, specifically the CDE Browser and Curation Tool, were used to map the NINDS core CDEs to the caDSR structure. This initial effort focused on the semantic interoperability between the NINDS CDE and caBIG® vocabularies. The mapping revealed that of 151 CDEs on 21 CRFs, 123 CDEs were identified as ''harmonized'' or partially harmonized with caDSR standards (81%). The successful mapping of the NINDS General CDEs serves as a proof of concept so that future NINDS disease-specific CDEs and CRFs can be integrated with the caBIG® standards. Furthermore, this effort demonstrates that even in different subject matter domains, such as neurology and oncology, a core of basic research variables can be created and shared among multiple groups. This reuse could permit thoughtful review of the remaining 20% of content that needs to be added to a metadata registry.

AUTHORS AND AFFILIATIONS
Yun Lu, PhD, KAI Research, Inc.; Dianne Reeves, RN, MSN, NCI CBIIT; Stacie Grinnon, MS, KAI Research, Inc.; Kristy Miller, MPH, KAI Research, Inc.; Patti Shugarts, BS, KAI Research, Inc.; Joanne Odenkirchen, MPH, NINDS

23. Bridging Basic Research and Its Translation: Integrated Preclinical and Clinical Science

For a decade, the NCI Center for Biomedical Informatics and Information Technology (NCI CBIIT) has collaborated with the NCI Mouse Models Consortium (NCI-MMHCC) to deploy resources that support the use of mice in cancer research. The maturing field of mouse model development provides opportunities to use the animals to inform translational and clinical research. The leading edge of this research is the NCI-MMHCC's Co-Clinical Trials project in which drug combinations are administered in parallel to genotyped patients, molecularly matched genetically engineered mice, and xenografted patient tissue explants. Specimens from the three cohorts are assayed similarly "for somatic mutations, germline SNP variations, response to specific regimens, imaging, microarray, and proteomics profiles." Then the integrated clinical, biological, and pharmacological information is used to identify genotypes and biomarkers that predict response to specific treatments. Additional NCI-MMHCC projects present opportunities to integrate mouse and human genetics to identify people at risk for cancer development or adverse reactions to early interventions or therapy. The multi-modal, cross-organism data correlations that are required in such studies present challenges for data aggregation, integration, and analysis that are beyond the scope of any single software tool. The caBIG® initiative is aimed at identifying and implementing modular software services that will support a standards-based approach to capturing and exchanging study data. The ability to recombine these services to support a variety of study designs will be critical to supporting this dynamic field.

AUTHORS AND AFFILIATIONS
Juli Klemm, PhD; National Cancer Institute Center for Biomedical Informatics and Information Technology; Cheryl Marks, PhD, National Cancer Institute Division of Cancer Biology

24. Visualizing and Exploring Genomic Alterations in the Cancer Genome Workbench

The Cancer Genome Workbench (http://cgwb.nci.nih.gov) provides a variety of tools for visualizing sample-level genomic data. Built originally on the UCSC genome browser, CGWB has extended the UCSC platform to include tracks for individual samples, as well as heatmaps, scatter plots, sequence traces, next-generation sequence alignments, and 3D protein structures. CGWB supports a number of high-throughput genomic data types, including copy number alterations, genotypes, somatic mutations, mRNA and miRNA expression, and methylation. The heatmap viewer offers interactive displays of gene expression and copy number changes along with clinical features. The next-generation sequence viewer, Bambino, allows users to examine the sequence alignment quality at base level and to identify SNPs/indels from next-generation sequence mapping files in SAM/BAM format. Data in CGWB is organized in projects and sub-projects. CGWB is currently hosting data from the following projects: The Cancer Genome Atlas (TCGA); Therapeutically Applicable Research to Generate Effective Treatments (TARGET); the Sanger Institute's Catalog of Somatic Mutations in Cancer (COSMIC); the NCI60 cell lines; and the GlaxoSmithKline cell lines. Much of the data in CGWB is available on caGrid. With comprehensive genomic alteration data from large numbers of tumor samples and cell lines, CGWB will help researchers gain new insight into cancer biology.

AUTHORS AND AFFILIATIONS
Chunhua Yan, Center for Bioinformatics and Information Technology, NCI CBIIT; Richard Finney, Center for Bioinformatics and Information Technology, NCI CBIIT; Michael Edmonson, Laboratory of Population Genetics, NCI; Carl Schaefer, Center for Bioinformatics and Information Technology, NCI CBIIT; Robert Clifford, Laboratory of Population Genetics, NCI; Cu Nguyen, Laboratory of Population Genetics, NCI; Shuang Cai, SAIC; Hongen Zhang, Laboratory of Population Genetics, NCI; Huaitian Liu, SAIC; Ying Hu, Laboratory of Population Genetics, NCI; Kenneth Buetow, Center for Bioinformatics and Information Technology, NCI CBIIT

25. Achieving Continuity of Care and Clinical Trials Success with caBIG® Clinical Trials Suite

Clinical trials are essential to advancing healthcare. However, they often present a challenge to continuity of care because patient healthcare information is segmented. The result of this segmentation hinders delivery of care and

presents obstacles to successful clinical trials outcomes. The Center for Biomedical Informatics and Information Technology at the National Cancer Institute has developed a comprehensive set of modular, interoperable and standards-based tools: the caBIG® Clinical Trials Suite for managing clinical data; and the Life Sciences Distribution Bundle for tracking and managing biospecimen data and analyzing and integrating genetic microarray data. These tools provide solutions to address segmentation issues. The new, service-based caBIG® Clinical Trials Suite facilitates data access, collaboration, standardization, and data integrity through the use of NCI Enterprise Services (NES). Suite application users can access a curated database of persons, organizations, and protocols through NES to assure data accuracy and consistency. Additionally, health care information collected prior to diagnosis can be retrieved seamlessly by the clinical trial site upon patient enrollment. Following enrollment into a clinical trial, the Suite is used for registration, scheduling, recording, and reporting adverse events and integrating lab information to maximize treatment and trial success. Following completion of the clinical trial, the patient's electronic medical record is updated and available to future care providers, ensuring continuity of care. Within the clinical trial framework, the ability to integrate data from the patient's profile, pathology, treatment, and outcomes allows clinical researchers to develop and refine strategies tailored to the individual's unique care needs. This poster describes a patient's health record following the patient from the US Department of Defense health system, into the caBIG® Clinical Trials Suite, then to a physician in a leading US Health Management Organization (Kaiser-Permanente) using the caBIG® data framework and the Nationwide Health Information Network.

AUTHORS AND AFFILIATIONS
Todd, Hardin, MS, National Cancer Institute Center for Biomedical Informatics and Information Technology; William, T., Dyer, Jr, Pyramed Research; John Speakman, National Cancer Institute Center for Biomedical Informatics and Information Technology

26. Facilitating Access to Biospecimen Research and SOPs through the Biospecimen Research Database

Pathology research and clinical practice relies on controlled procedures to minimize specimen processing artifacts to ensure accurate analysis and diagnosis. There is a need to identify research on specimen handling in peer-reviewed journals and to identify standard operating procedures (SOPs) for broad application across multiple institutions. This will enable pathologists to validate research findings and lead to more rapid transition of translational research into laboratory and clinical practice. The National Cancer Institute's Biospecimen Research Database (BRD) provides pathology practitioners and researchers an annotated source of published research and review articles of specific relevance to human quality biospecimen collection and processing; it can also be accessed at http://biospecimens.cancer.gov/brd/. Over 300 papers have been indexed according to test analyte, specimen type and experimental variables. Additional functionality includes curation tools and experimental factor search options. An additional 700 publications have been identified for curation. Future functionality includes opportunities for community comment, meta-analysis of papers, and addition of electronic SOPs. The BRD will become a definitive source of structured SOPs that can be searched individually, yet can also be used by software systems (i.e. cancer Bioinformatics Grid-caBIG® applications and others). SOPs will exist as structured data that can be queried in caGrid-enabled data services. The BRD is intended to be used by pathologists, researchers and biorepository managers who wish to search, subscribe, and comment on SOPs and the supporting annotated published literature. OBBR's goal is for the BRD to serve the needs of the biospecimen science and pathology communities. Feedback and paper referrals are welcome and can be submitted at biospecimens@mail.nih.gov.

AUTHORS AND AFFILIATIONS
Andrew W. Breychak, BA, NCI CBIIT, Sapient Government Services; Ian Fore, DPhil, NCI CBIIT; Anna Fernandez, PhD, NCI CBIIT, Booz, Allen Hamilton; Helen Moore, Phd, NCI-Office of Biorepositories and Biospecimen Research; Kelly B. Engel, PhD, NCI-Office of Biorepositories and Biospecimen Research, Preferred Staffing Services

27. NCI Terminology Browsers Offer New EVS Content and Services

NCI's new terminology browsers provide a user friendly interface to all terminologies supported by NCI Enterprise Vocabulary Services (EVS). Key improvements have been made in completeness, consistency, cross-links, user suggestions, and performance. Work has begun on new mapping and subset functionality tied to LexEVS 6.0 terminology server extensions, and advanced search and other new features are being added. The NCI Term Browser is a Web application that gives access to all biomedical terminologies hosted by EVS. It extended the NCI Thesaurus (NCIt) Browser to some 20 additional terminologies, supporting search and cross-links with the NCI Metathesaurus (NCIm) Browser that itself offers 76 terminologies cross-mapped to capture shared meanings. Each Term Browser terminology has its own home page and can be searched and used separately or together. Users can search by name/id, property, or relationship values using exact match, begins with, or contains algorithms. The user may browse terms, codes, definitions, relationships, and other information for any concept, or navigate concept hierarchies. NCIt continues to play a central role in the NCI/caBIG® semantic infrastructure and information services, including over 15,000 terms in some 150 subsets maintained with partner organizations including the Food and Drug Administration (FDA), the Clinical Data Interchange Standards Consortium (CDISC), and the National Council for Prescription Drug Programs (NCPDP). LexEVS 6.0 support for the HL7/OMG Common Terminology Services 2 (CTS 2) specification will help make these easily accessible through both the server and the browser, together with many other subsets not currently supported. Mappings between terminologies and terminology subsets will also be much better supported with CTS 2. EVS has often provided such mappings based on NCIm, or on term or code matching between terminologies. CTS 2 will support formal representation and storage of such mappings; moreover, the browsers are being extended to make them accessible to users.

AUTHORS AND AFFILIATIONS
Lawrence Wright, MA, National Cancer Institute Center for Biomedical Informatics and Information Technology; Wilberto Garcia, MS, Northrop Grumman, Steven Hunter Associates, Ekagra; Jason Lucas, BS, Northrop Grumman; Kim Ong, PhD, Northrop Grumman; Tracy Safran, BS, SAIC; Frederick Robert Wynne, Lockheed Martin; David Yee, MS, Northrop Grumman

28. caIntegrator2: A Translational Research Tool for 21st Century Biomedicine

.

caIntegrator2 is a Web-based software package that allows researchers to set up custom, caBIG®-compatible Web portals to conduct integrative research without requiring programming experience. These portals bring together heterogeneous clinical, microarray and medical imaging data to enrich multidisciplinary research. Using caIntegrator2, researchers can execute, save and share queries to identify and collect many types of data, combining clinical information with genetic and genomic data to enable multidimensional analysis. Users can also run advanced queries and perform correlative outcome analysis using Kaplan-Meier plots, gene-expression plots, and other tools. caintegrator2 uses caGrid analytical services such as GenePattern and BioConductor to perform analyses on the integrated study data, including copy number analysis. caIntegrator2 leverages the Cancer Data Standards Registry and Repository (caDSR) to map experimental data to well-defined datatypes and utilizes caGrid and Java client APIs to access data from caBIG® applications such as caArray, the National Biomedical Imaging Archive (NBIA). caIntegrator is also integrated with caBIO (Cancer Bioinformatics Infrastructure Objects) to perform queries on genes and pathways. Data access in caIntegrator2 is controlled by role-based authorization, permitting portal data to be either public or private. caIntegrator2 is an open-source, Web-based J2EE Java application employing Spring, Struts 2 and Hibernate technologies.

AUTHORS AND AFFILIATIONS
Mervi Heiskanen, PhD, NCI CBIIT; Annad Basu, MS, MBA, NCI CBIIT; Juli Klemm, PhD, NCI CBIIT; JP Marple, MS, 5AM Solutions; Huaitian Liu, PhD, SAIC; Ngoc Nguyen, MS, Claris; TJ Andrews, MS, ScenPro; Karen Ketchum, PhD, ESAC Inc.; Quy Phung, PhD, ESAC Inc.; Jacob Shine, MS, ESAC Inc

29. caLIMS2: A Next Generation Cancer Laboratory Information Management System Designed to be caBIG® Interoperable

The purpose of the caLIMS2 project is to create a Laboratory Information Management System (LIMS) that is interoperable within established caBIG® standards and guidelines. Currently, caBIG® tools exist for Clinical Trials implementation and monitoring; biospecimen storage and tracking; data storage and retrieval; and data analysis. However, there exists a gap in the caBIG® tool set related to the capture and tracking of laboratory activities. caLIMS2 will fill that gap – tracking a complete laboratory workflow that uses materials from a specimen management service (e.g. caTissue) to generate experimental results for one of the caBIG® data management services (e.g. caArray). Core caLIMS2 functions are organized in four basic modules (administration, inventory, workflow, and reports) and include the management of personnel, equipment, lab supplies and reagents, samples, laboratory workflow and experimentally derived metadata and data. caLIMS2 is highly flexible, making it suitable for research labs and high throughput core facilities and for support of multiple research domains (genomics, proteomics, nanoparticle characterization, etc.). caLIMS2 will help further translational cancer research through the organization of laboratory workflow; tracking of specimens and derived specimens; acquisition of laboratory data and metadata associated with those specimens; and the appropriate sharing and dissemination of the data to support subsequent analyses. caLIMS2 v1.0 focuses on basic functionality and easy integration of caBIG® specimen management and data management services. We will present the caLIMS2 v1.0 UML model, a prerelease package, and describe our progress in developing the version 1.0 release. caLIMS2 v1.0 includes design for integration with other caBIG® application services (e.g. caArray and caTissue) and the creation of two enterprise level services: Equipment Service and Experiment Method Service. More details and information on the project can be found on the caLIMS2 Project Wiki site (https://wiki.nci.nih.gov/x/2oMYAQ) and the caLIMS2 GForge site (http://gforge.nci.nih.gov/projects/calims2/).

AUTHORS AND AFFILIATIONS
Jenny Kelley, MA, NCI/CCR/LPG; Robert Clifford, PhD, NCI/CCR/LPG; Sashi Thangaraj, MS, Moxie Informatics, Inc., Cu Nguyen, BS, NCI/CCR/LPG; Henry Zhang, PhD, NCI/CCR/LPG; Anand Basu, MS, NCI/OD/CBIIT; Kenneth Buetow, PhD, NCI/OD/CBIIT

30. Grid-Enabled Measures (GEM) Database: An Informatics Tool to Support Cancer Prevention and Control Research

In this era of “Big Data,” it is crucial to develop tools that facilitate the use of standardized measures and allow for sharing the resulting harmonized data via an interoperable platform. This presentation will describe the development and implementation by the National Cancer Institute of the Grid-Enabled Measures (GEM) tool for researchers. GEM is a Web portal mounted on a grid cyberinfrastructure (internet-based research environment) – that allows researchers to search for measures based on theoretically-based constructs (the concept the measure assesses); examine metadata about these measures (e.g., validity, reliability, constructs, history, and usage); and download the measures as PDF files and submit new ones. Users can also give feedback about existing constructs and measures using ratings and comments and can interact with the virtual community of users through a Wiki-like process. In addition, this cyberinfrastructure-based tool allows users to upload and exchange data with common data elements to promote data harmonization. The poster will show how GEM contains many constructs and associated measures that are directly applicable to studying cancer prevention and control such as risk behaviors (e.g., tobacco dependence, diet and nutrition, physical activity, and cancer screening) and other cancer-related factors (e.g., health literacy, numeracy, quality of life, and depression). Lastly, there will be a section on the challenges of deciding on criteria for vetting and approving the measures, and for creating a viable ''business model'' to provide incentives to the research community to agree upon the use of common measures and sharing of data.

AUTHORS AND AFFILIATIONS
Richard, Moser, PhD, Behavioral Research Program, NCI/Division of Cancer Control and Population Sciences; Bradford W. Hesse, PhD, NCI/Health Communication and Informatics Research Branch; Abdul, R. Shaikh, PhD, MHSc, NCI/Behavioral Research Program, Division of Cancer Control and Population Sciences; Paul, Courtney, MS, SAIC supporting Division of Cancer Control and Population Sciences/NCI; Gordon Willis, PhD, NCI/Applied Research Program, Division of Cancer Control and Population Sciences; Erik Augustson, PhD, MPH, NCI/Tobacco Control

Research Branch, Division of Cancer Control and Population Sciences; Kerry, Levin, PhD, Westat, Inc.; Cynthia Helba, PhD, Westat, Inc.; Kisha Coa, MPH, Westat, Inc.; David, Garner, Westat, Inc.; Marsha, Dunn, MPH, Health Studies Sector, Westat, Inc.

31. Global Standards for Grading Prostate Cancer

Reliable diagnosis of prostate cancer severity is a challenge in clinical practice. The reported rate of reproducibility with respect to the grading of prostate cancer by different pathologists hovers at a mere 30-60% (T. V. d. Kwast et al, BJU International, 2004). Current methodology suffers from systematic and non-systematic errors, as well as individual scorers' bias and inconsistency. In addition, digitization of existing slide collections and increasing prevalence of digital microscopes is changing the landscape of cancer research. Thus there is strong impetus to develop reliable automatic image processing techniques to supplement manual diagnosis. In our work, biopsy samples are stained with hematoxylin and eosin (H&E) and organized into tissue microarrays (TMA). We image the TMA cores and apply a pattern recognition approach for automatic analysis by computing a large generalized set of image features to describe each image. Additionally, we use alternate representations of the pixel plane via image transforms to extend the range of the image content considered (N. Orlov et al, PRL, 29, 2008). This approach was successfully applied to classification of lymphoma types (N. Orlov et al, IEEE Tr. IT Biomed, accepted) as well as the stages of melanoma progression (N. Orlov et al, Conf. Vis., Imaging and Image Proc., Palma de Mallorca, Spain, 2008). Preliminary results for automatic Gleason grading of prostate cancer attained per-image accuracy of 90%. Several advantages of computer-based prostate grading include the ability to assign grades to images automatically as well as provide consistency and standardization of diagnosis. Computer classifiers could be trained based on terminal outcome data instead of Gleason scores to provide a more direct prognosis tool. These studies are ongoing as terminal outcome data become available.

AUTHORS AND AFFILIATIONS
Nikita V. Orlov, PhD, NIA/NIH; Chris Coletta, NIA/NIH; John Delaney, PhD, NIA/NIH; D. Mark Eckley, PhD, NIA/NIH; Salim Rahimi, NIA/NIH; Lior Shamir, PhD, NIA/NIH; Ilya G. Goldberg, PhD, NIA/NIH

32. XNAT Neuroimaging Database of GCAP (XNAT@GCAP)

XNAT@GCAP is a neuroimaging database adapted from the early version of XNAT (The Extensible Neuroimaging Archive Toolkit, a BIRN sponsored project) to meet the brain imaging research requirements of the Genes, Cognition and Psychosis Program (GCAP) of NIMH. It was customized to manage and optimize data archiving, data de-identification, data processing, results inspection, and data sharing of neuroimaging data. It also facilitates data mining and quality control procedures. At this point all these procedures have been highly automated for fMRI and structural MRI data. The data are downloaded automatically from the NIH MRI Research Facility DICOM image server every day, de-identified, archived and prepared in the storage system for further analysis. Data processing parameters are managed in the database for easy review and modification by users. Large scale data processing is made possible as the data processing module can distribute the data processing tasks into a cluster of Linux computers. Users can initialize and monitor the progress of the data processing, and optionally receive email notifications. Along with image pre-processing, some quality control (QC) statistics are also generated. These fMRI and sMRI statistical results (e.g., first-level individual subject maps, QC measures, segmentation images, etc.) are available on-line for users to perform visual inspection and provide subjective QC remarks. The database can then be used to dynamically filter the existing results based on thresholds set for the quantitative QC measures and subjective QC remarks. In addition to fMRI and sMRI, the database can also manage MR spectroscopy data. The data archiving, de-identification and processing steps for these modalities are also partly automated. A Web interface is also available for the user to inspect 3-D high resolution sMRI images online. Currently about 2600 fMRI datasets, 2000 structural MRI datasets and 250 spectroscopy datasets have been archived in this database.

AUTHORS AND AFFILIATIONS
Xi Cheng, NIMH; Daniel R. Weinberger, NIMH/NIH; Daniel Marcus, NIMH/NIH: Fabio Sambataro, NIMH/NIH; Brad Zoltick, NIMH/NIH; Yunxia Tong, NIMH/NIH; Andreas Meyer-Lindenberg, NIMH/NIH; Venkata S. Mattay,

Neuroimaging Core Facility, Genes, Cognition and Psychosis Program, NIMH/NIH, Mallinckrodt Institute of Radiology, Washington University

33. Biospecimen Reporting for Improved Study Quality (BRISQ)

The amount of detail reported concerning biospecimen characteristics and handling varies widely in scientific publications. We addressed this by compiling reporting recommendations for biospecimens. A workshop was held at the 2009 Biospecimen Research Network Symposium to form biospecimen reporting recommendations. The resultant Biospecimen Reporting for Improved Study Quality (BRISQ) list was refined at monthly teleconferences by a committee of workshop attendees and other experts. The BRISQ committee composed a three-tiered list of biospecimen data that should be reported, if known and applicable, for all biospecimens or patients in the study. When consulting the BRISQ list, researchers should evaluate and adjust the weight of each item in the context of the study. The first tier, items necessary to report, includes biospecimen type; relevant clinical characteristics; collection, stabilization, and preservation mechanism; type and composition of long-term preservative; storage temperature and duration; shipping temperature(s); and composition assessment and selection. Items advisable to report form the second tier: patient demographics; accrual scheme; time and temperature between acquisition and stabilization; type of collection container; time and temperature in preservation solution; aliquot volume; shipping duration(s); gross and microscopic review; proximity to relevant anatomical lesion; and details of enrichment for relevant component(s). Additional items to report, or third-tier items, include agonal state and cause of death for postmortem biospecimens; relevant exposures; reproductive status; nature of biobank; time from blood flow cessation to acquisition; specimen size; type of storage container/slide and shipping vessel; shipping conditions; number, time, and temperature of any of freeze-thaw events; embedding medium; and quality assurance measures. If followed, these recommendations will provide readers with sufficient information to evaluate, interpret, compare, and reproduce the results of experiments that employ human specimens. Thus, not only might overall quality of publications improve, but the quality of biospecimen investigation might improve over time.

AUTHORS AND AFFILIATIONS
Andrea, Kelly, PhD, Rose Li and Associates, Inc.; Scott Jewell, PhD, The Ohio State University Department of Pathology and Comprehensive Cancer Center, Biorepository and Biospecimen Resource; Lisa McShane, PhD, NCI/Biometric Research Branch; Helen M., Moore, PhD, Office of Biorepositories and Biospecimen Research/NCI; Jim, Vaught, PhD, NCI/Office of Biorepositories and Biospecimen Research, Members of the BRISQ Committee

34. Bristol-Myers Squibb (BMS)/National Cancer Institute (NCI) Pilot

The proposed abstract explores how BMS and NCI's Cancer Therapy Evaluation Program (CTEP) and SAFE-BioPharma are collaborating to demonstrate how an all electronic workflow speeds up business process flow and improves NCI's ability to accelerate research and be more responsive to public health. BMS directly and indirectly communicates with many organizations in NCI. Current methods include Federated Identify Management. Federal Bridge PKI methods further enhance NCI's capabilities. Integrating these technologies with existing processes and digital signatures for approvals will improve business processes and will reduce the time required for trials and quicken the time to market. The pilot, which at this writing is underway, is expected to provide a model for broader use of Federal Bridge and SAFE-BioPharma compliant digital signatures in business process flows within NCI and NIH. The pilot has three primary components: (1) Cross certified credentials – three different types of credentials from 3 groups (BMS staff; NCI personnel and Cooperative Groups) will be used for digitally signing documents; (2) Types of Documents – documents being signed include, but are not limited to, Letter of Intent, Clinical Trial Agreements, contracts, Concept Approval and Protocol Approval; (3) Workflow – the pilot is configured to alert the signatory that a document needs to be signed and to alert the document owner when all signatures have been obtained. Since the BMS/NCI digital signing pilot eliminates the need for wet signatures and therefore, paper, it improves the speed and efficiency by which medical research is conducted. To put this into context, the New England Journal of Medicine has estimated that 40% of the cost of bringing a new drug to market is tied to paper-based processes.

AUTHORS AND AFFILIATIONS
Tanya Newton, Manager, Regulatory Affairs & Compliance Operations, SAFE-BioPharma Association; Jon Weisberg, Director, Communications and Public Relations, SAFE-BioPharma Association; Cindy Cullen, Chief Technology Officer, SAFE-BioPharma Association

35. Developing a Collaborative Environment Supporting the Application of Nanotechnology in Biomedicine

The application of nanotechnology in cancer promises advancements in early detection, targeted therapeutics, and cancer prevention and control. The use of nanotechnology in biomedicine involves the engineering of nanomaterials to act as therapeutic carriers, targeting agents, and diagnostic imaging devices. To assist in expediting and validating the use of nanomaterials in biomedicine, the NCI Center for Biomedical Informatics and Information Technology (CBIIT), in collaboration with the NCI Nanotechnology Characterization Laboratory (NCL) and other Cancer Centers of Nanotechnology Excellence (CCNEs), has developed a data sharing portal called caNanoLab. caNanoLab provides access to experimental and literature curated data from the NCL, Washington University, and other CCNEs. caNanoLab facilitates data sharing via the use of caBIG® technologies (caGrid), enabling semantic interoperability and data exchange between other caBIG® tools. caNanoLab is based on a nanotechnology object model (nano-OM) which acts as a standard representation of nanomaterials and their physical (e.g. size, molecular weight) and in vitro (e.g. cytotoxicity, immunotoxicity) characterizations. The nano-OM leverages and extends concepts from the NCI's Enterprise Vocabulary Services (EVS) and the Nanomaterial Ontology (NPO) designed by Washington University. The nano-OM provides a model for representing the composition of diverse nanomaterial types (e.g. dendrimer, fullerene, quantum dot, carbon nanotube) and associated functionalizing entities (small molecules, antibodies). These functionalizing entities allow particles to achieve the desirable therapeutic or diagnostic functions and enable personalized medicine via the administration of targeted therapies. The project is expanding to include support for in vivo characterizations of nanomaterials and their functionalizing entities, which are analogous to those required for small molecules and other medical devices. These characterizations involve rigorous testing to determine toxicity and pharmacokinetics properties. The caNanoLab project is collaborating with members of the biomedical nanotechnology community through the caBIG® Nano WG in the development of nano-TAB, a standard supporting data import/export between disparate nanotechnology systems.

AUTHORS AND AFFILIATIONS
Juli Klemm, PhD, NCI CBIIT; Anand Basu, NCI CBIIT; Piotr Grodzinski, PhD, NCI OTIR; Krzysztof Ptak, PhD, NCI OTIR; Anil Patri, PhD, NCL; Marty Fritts, PhD, NCL; Sharon Gaheen, MBA, SAIC; Sue Pan, SAIC; Thai Le, SAIC; Elizabeth Hahn-Dantona, PhD, Lockheed

36. caGrid Portal: caGrid Information Visualization

The caGrid Portal is a Web-based application that enables you to discover and interact with the services that are available on the caGrid infrastructure. The portal serves as the primary visualization tool for the caGrid middleware and provides a standards-based platform for hosting caBIG®-related tools. It also serves as a caBIG® information source. Using the caGrid portal, you have instant access to information about caBIG® participants, caGrid points of contact (POCs), and caGrid-related news and events. caGrid Portal 3.0 also allows to share content like caBIG® tools, services, and caGrid data service queries with other users of the caBIG® community. caGrid Portal also allows for the creation of ad-hoc communities within the portal that can be administered and managed by Portal users.

AUTHORS AND AFFILIATIONS
Joshua Phillips, NCI CBIIT; Manav Kher, Semantic Bits

37. The Cancer Gene Index Project

An accurate and up-to-date inventory of cancer genes is a necessary foundation for advancing cancer research and supporting patient treatment. Clinical researchers and MDs treating patients all easy access to a reliable core knowledge repository in order to connect “omics” with “oligies.” In 2004, NCI launched the Cancer Gene Index Project to provide the cancer community with a complete compendium of all cancer related genes occurring in the biomedical literature with manually annotated gene/disease and gene/compound relationships. The aim was to accelerate discovery of cancer drugs, biomarkers, cures and treatments. A cancer gene was defined as any human gene or gene product that co-occurs in a single Medline sentence with a cancer disease or compound/treatment term. The Biomax BioLT Linguistics Tool was used to automatically analyze the complete Medline database with more than 18M abstracts. During the 5-year project, 6,955 cancer genes were identified and 1.8M sentences were manually validated and annotated with role codes and evidence codes for each gene/disease and gene/compound/treatment relationship. Sophic has integrated the Cancer Gene Index into the Biomax BioXM Knowledge Management Environment to support research at NCI’s Center for Cancer Research. BioXM was the first caBIG® Bronze compliant commercial system and is configurable to Silver Compatibility. BioXM and Cancer Gene Index are configured to support translational medicine, biomarker, mRNA and pathway discovery use cases. The complete Cancer Gene Index data are available without restriction to the community through the National Cancer Institute’s caBIO gene annotation database (https://wiki.nci.nih.gov/display/ICR/caBIO) or via bulk FTP download (http://ncicb.nci.nih.gov/NCICB/projects/cgdcp). Sophic also provides access to the Cancer Gene Index and two other databases – the Sophic Druggable Genome and The Genome Atlas (Brain, Ovarian and Lung Cancer Genes) at www.sophicalliance.com.

AUTHORS AND AFFILIATIONS
K. Albermann, Biomax Informatics AG; A. Fritz, Biomax Informatics AG; K. Wenger, Biomax Informatics AG; K. Heumann, Biomax Informatics AG; G.A. Komatsoulis, National Cancer Institute Center for Biomedical Informatics and Information Technology; J.D. Klemm, National Cancer Institute Center for Biomedical Informatics and Information Technology; P.M. Blake, Sophic Systems Alliance, Inc.

38. The TRIAD Project: Adopting and Adapting caGrid for the CTSA Environment

The availability of scalable, extensible, service-oriented biomedical informatics platforms is critical to the performance of efficient, timely, and high quality research in multi-site or investigator settings. Such requirements are further emphasized when research is being conducted across the full translational spectrum involving a broad variety of information sources, stakeholders, and analytic resources. We report upon the design, initial adoption, and future plans for such an informatics platform, as part of the activities of The Ohio State University’s CTSA-funded Center for Clinical and Translational Science, which is known as the Translational Informatics and Data-management (TRIAD) project. As part of the TRIAD, we have deployed and extended the caGrid middleware as our core data and knowledge-sharing platform. Specific extensions to caGrid have included: 1) the implementation of OpenMDR, an extensible knowledge management component that extends the UK’s CancerGrid metadata repository (cgMDR) platform, which is an open standards and open source implementation of the ISO metadata registries standard (ISO11179-3) and allows users to define, manage, and utilize locally relevant metadata element definitions as well as access standard metadata repositories; 2) the deployment of software components capable of linking the GAARDS security system with institution-level authentication systems; and 3) the design, evaluation, and integration of software components capable of publishing or consuming information contained in common research data management systems, such as i2b2 and REDCap. At the time of this submission, TRIAD has been deployed for use at OSU, and is actively being utilized in multiple projects, including: 1) longitudinal tracking of phenotypic data for maternal-fetal dyads in support of perinatal outcomes research; 2) discovery of patient and tissue cohorts spanning numerous institutional bio-specimen repositories and an enterprise data warehouse; and 3) development of an integrative clinical trial and bio-specimen management solution for multi-center oncology research.

AUTHORS AND AFFILIATIONS

Philip R.O. Payne, PhD, The Ohio State University; Justin Permar, The Ohio State University; David Ervin, The Ohio State University; Rakesh Dhaval, MS, The Ohio State University; Calixto Melean, MS The Ohio State University, ; Tara B. Borlawsky, MA, The Ohio State University

39. Structure through Minimum Information Guidelines: Information-Focused Approach to Scientific Publishing

We would like to submit a poster on the development of an e-journal for publishing therapy experiments data. The approach to this e-journal is unique in that our publications will be based on experiment data recorded according to the minimum information guidelines called GIATE (Guidelines and Information About Therapy Experiments). Consequently, our journal places the focus on sets of GIATE-required information and data, while supplementing them with descriptions for context and publications for background. GIATE requires information about the molecular target, therapeutic agent and models. We use this structure as the “Table of contents” for each publication. The purpose behind this approach is to collate a consistent set of information regarding these aspects of therapy experiments. In this poster we will address the matter of information resolution. We want to find the minimum resolution of experiment details that provide enough clarification of experiments details. This leads to the issue of natural prose language versus the usage of terms from controlled vocabularies and ontologies. Scientific information is traditionally presented in the natural prose language of scientific publications. Freedom of expression provides the nuance necessary for the effective dissemination of knowledge. Articles can be integrated with ontologies using tools which auto-annotate specific terms with meanings. However, the specificity of the information provided differs according to providers. We will provide a case study of information recorded when a user is allowed free text to describe a fusion protein as a therapeutic agent (e.g., it's molecular target, components, production system, mechanism of action, etc.), and contrast it with the information recorded when the user is required to provide the same information using terminology from NCI Thesaurus.

AUTHORS AND AFFILIATIONS
May Yong, PhD, University College London ; Richard Begent, Professor of Clinical Oncology, University College London

40. PALMS-CI: A Policy-Driven Cyberinfrastructure for the Exposure Biology Community

The objective of the Exposure Biology program is to develop tools that assist researchers in understanding the relationships between assessments of behavior and exposure to health-related factors in the environment. Physical activity is a behavior that has been poorly measured in exposure biology research. To address this issue the EB program has invested in the development of wearable devices that measure location and motion, leaving much of the data management and analysis for future work. At the same time, the NSF has required that researchers make their data available for use by other researchers, and HIPAA regulations place significant restrictions on data sharing. Finally, as personalized sensors proliferate, significant opportunities arise for large scale and continuous collection, analysis, and leveraging of personal data. The PALMS (Physical Activity Location Measurement System) project has developed a cyberinfrastructure (PALMS-CI) as a highly scalable, policy-driven system capable of managing organizational and observational data for a number of studies and ultimately facilitating NSF, HIPAA, and other policy objectives. Researchers define the data maintained for their studies, the data collection devices, and the particular protocols used to analyze and visualize the results using packages such as SPSS and ArcGIS. In providing a consistent and common infrastructure, the PALMS-CI enables policy-based sharing and reuse of data and analytics thereby enabling and incentivizing the participation of myriad stakeholders including researchers, funding organizations, and regulatory agencies. The PALMS-CI leverages the Rich Services variant of Service-Oriented Architecture (SOA) techniques realized on an Enterprise Service Bus (ESB), resulting in exceptional flexibility in accommodating new and changing requirements. Consequently, PALMS-CI is a research platform for emerging technologies such as policy design, decision, and evaluation frameworks; dynamic service binding; service-oriented modeling; transaction modeling in loosely coupled systems, storage virtualization; service-oriented validation and verification; large scale data acquisition cloud computing; and research-oriented social networks.

41. caBIG® Workflow Infrastructure using Taverna

caBIG® provides a variety of data and analytical services targeted towards Biomedical research. Some of the tasks in the collection and analysis of cancer-related data for answering specific Biological or Biomedical questions require orchestrating multiple caGrid Data and Analytical services as workflows. For the empowerment of users from biological or biomedical domains in creating and executing their workflows efficiently, the caBIG® Workflow Team has selected the Taverna workbench and has successfully created a workflow infrastructure to orchestrate caGrid Data and Analytical services. This caGrid Workflow Infrastructure aims at providing: (1) Easy-to-use workflow authoring and submission tool that allows semantic based service discovery, Grid security configuration and enforcement; support for stateful Grid services and many other caGrid specific features in form of Taverna plugins (e.g., cagrid-discoverer, workflow-execution-service, cagrid-activity, and cagrid-transfer); (2) Taverna Workflow Execution Service that enables users to submit workflows created within Taverna workbench to this WSRF service. It allows asynchronous submission of workflows and provides a comprehensive client API to mange the workflows submitted. It is caGrid security compliant and allows clients to delegate their credentials to the service to access secure services involved in the workflows. It also supports caGrid Transfer service that allows http transfers of large input and output files involved in the workflows; and (3) Taverna Execution Portlet, a Web-based interface that allows users to discover existing workflows and submit them to the Workflow Execution Service along with the input data.

AUTHORS AND AFFILIATIONS
Dinanath Sulakhe, MS, Computation Institute, University of Chicago; Wei Tan, PhD, Computation Institute, University of Chicago; Ravi Madduri, MS, Computation Institute, University of Chicago and Argonne National Laboratory; Stian Soiland-Reyes, MS, University of Manchester; Alexandra Nenadic, PhD, University of Manchester

42. Integrative Computational Platform for the Systems Biology of Adenoid Cystic Carcinoma

Systems biology is often faced with the daunting task of integrating and analyzing high-throughput experimental data. In order to generate biological insight from these data, it is important to systematically integrate disparate high-throughput data types. Furthermore, the integration of various data for specific pathologies is limited. Adenoid Cystic Carcinoma (ACC) is a cancer that starts from glandular tissues and spreads along the nerves. With support from the Adenoid Cystic Carcinoma Research Foundation (ACCRF), multiple research groups have generated extensive data profiling human patient gene expression, gene and protein expression in xenograft models, and cell line responses to various pharmaceuticals, among other rich data sources. Here, we employ MySQL database technology to store ACC xenograft data using a new schema design and allowing access under a single data platform. This schema can accommodate various features representative of transcriptomic and proteomic data for a variety of conditions. We also developed a Web interface (http://accdb.bme.virginia.edu/) that provides researchers with integrated information for different types of high-throughput data specific to ACC tumors. To demonstrate the utility of a disease-focused computational platform integrating disparate data, we present several case studies such as the analysis of drug sensitivity based on changes in gene expression values and the correlated effect on downstream pathways due to the targeting of receptor tyrosine kinases.

AUTHORS AND AFFILIATIONS
Sudhir Chowbina, MS, University of Virginia; Christopher Moskaluk, MD, PhD, University of Virginia; Jason Papin, PhD, University of Virginia

43. An Open-Source Integrative Framework for Biomedical Image Analysis

Numerous, complementary software packages are available for large-scale medical imaging data analysis. Typically, these frameworks are not readily compatible in terms of workflows or data formats, so image scientists and clinical investigators are constrained to collaborate using a single package. Often, these constraints drive

researchers to choose a package that is less natural for the problem at hand for any individual investigator, and, instead, favor a compromise package. Furthermore, algorithms with a wide variety of applications are implemented in only one package, and must be reinvented, re-implemented, or accessed via complicated scripting techniques to be taken advantage of in other packages. The complications stemming from software package dependence impede innovation and collaboration amongst researchers by impeding dissemination. We propose a compromise between the independent development of packages and a single uniform collaboration framework through automated framework integration. This approach enables development and code reuse within a variety of software packages –each of which may lend itself naturally to specific applications for specific developer-scientists. Therefore, algorithms may be easily explored in new contexts and used within other software tools. Here we demonstrate an open-source integration effort across four medical imaging packages: Medical Image Analysis and Visualization (MIPAV), Java Image Science Toolkit (JIST), command line tools, and 3D Slicer. We explore three case studies: (1) a system for MIPAV to fully expose internal algorithms; (2) a method for self-documenting command line programs which provide for inclusion of JIST algorithms in scripting environments; and (3) an eXtensible Markup Language (XML) approach which allows 3D Slicer to detect and utilize JIST modules. We review the challenges and opportunities for light-weight software integration both within development language (e.g., Java in MIPAV and JIST) and across languages (e.g., C/C++ in 3D Slicer and shell in command line tools).

AUTHORS AND AFFILIATIONS
Bennett A. Landman, PhD, Vanderbilt University; Kelsie Covington, Vanderbilt University; Evan S. McCreedy, PhD, National Institutes of Health; Min Chen, BS, Johns Hopkins University; Aaron Carass, MS, Johns Hopkins University; Nicole Aucoin, BSc, Brigham and Women's Hospital; Jerry Prince, PhD, Johns Hopkins University; Dzung Pham, PhD, Center for Neuroscience and Regenerative Medicine

44. Somatic Mutations, Interoperability, Integration and Data Standards

Cancer is a genetic disease affecting approximately one in three individuals in Europe and North America. The Cancer Genome Project (CGP) was started at the Wellcome Trust Sanger Institute in 2000 to identify somatic changes and genes implicated in cancer. As part of this project a database, the Catalogue of Somatic Mutations In Cancer (COSMIC) was developed to store mutations uncovered by the project and those curated from the scientific literature (http://www.sanger.ac.uk/genetics/CGP). In order to maximize the value of this resource we have standardized the data enabling us to integrate and interoperate with external data resources. For example, we have standardized the description of somatic mutations using HGVS nomenclature and made the data available through the Distributed Annotation System (DAS). This Web service enables external resources to integrate information from multiple distant servers and present them in a single consolidated view. This enables our somatic mutation data to be easily viewed in ensemble, allowing the user to see the genomic context of the mutation and in Pfam, to explore the protein context of these mutations. In addition to looking into the context of a mutation we provide a simple and rapid method for mining the data. This is achieved using a query orientated data management system, BioMart (http://www.sanger.ac.uk/genetics/CGP/cosmic/biomart/martview). This allows the user to rapidly mine through complex biological data using only a simple Web interface. BioMart is being used to hold the International Cancer Genome Consortium (ICGC) data. The ICGC is a project to sequence and characterize 500 tumors and normals from 50 different tumor types. This data will include expression, copy number and full genome sequence for all samples. As COSMIC and the ICGC franchise database make the data available through BioMart using similar data models the resources will be fully interoperable.

AUTHORS AND AFFILIATIONS
Adam Butler, The Wellcome Trust Sanger Institute; Jon Teague, The Wellcome Trust Sanger Institute; Simon Forbes, The Wellcome Trust Sanger Institute; Dave Beare, The Wellcome Trust Sanger Institute; Nidhi Bindal, The Wellcome Trust Sanger Institute; Kenric Leung, The Wellcome Trust Sanger Institute; Rebecca Sheperd, The Wellcome Trust Sanger Institute; Sally Bamford, The Wellcome Trust Sanger Institute; Charlotte Dunham, The Wellcome Trust Sanger Institute; Andy Menzies, The Wellcome Trust Sanger Institute; Dave Jones, The Wellcome Trust Sanger Institute; Keiran Raine, The Wellcome Trust Sanger Institute; John Marshall, The Wellcome Trust Sanger Institute; Lucy Stebbings, The Wellcome Trust Sanger Institute; Chai Yin Kok, The Wellcome Trust Sanger

Institute; Mingming Jai, The Wellcome Trust Sanger Institute, Peter Campbell, The Wellcome Trust Sanger Institute; P Andrew Futreal, The Wellcome Trust Sanger Institute; Michael R Stratton, The Wellcome Trust Sanger Institute

45. COSMIC Database and Standards for Somatic Mutations implicated in Human Cancer

Cancer is a complex disease caused by an array of genetic alterations to include point, insertion, deletion, structural rearrangement and copy number changes, the majority of which are thought to be somatic and have accumulated during the lifetime of the individual tumor. A vast amount of information has accumulated on these alterations. However, most of these data are dispersed across the scientific literature with very little standardization. This has meant using and mining the data is extremely difficult. To address this situation, we have developed the Catalogue of Somatic Mutations In Cancer (COSMIC), which was released in 2004, to curate the scientific literature for somatic mutations (http://www.sanger.ac.uk/cosmic). The majority of point mutated cancer genes have now been curated and the database has been successfully adapted to curate cancer gene fusions. Release 46 contains information on 115,737 mutations, 449,676 tumors and 8,911 publications. With the advent of Next Generation Sequencing and the anticipated increase in mutation data, COSMIC has been adapted to handle data from these technologies. In order to maintain the data quality in COSMIC and allow interoperability with other data resources the project makes use of existing data standards. For instance, the database uses HGVS mutation nomenclature to describe each mutation. The group has also been actively involved in developing data models to represent somatic mutations with other cancer databases, the NCRI Informatics Initiative and more recently the International Cancer Genome Consortium project (ICGC). If we are to maximize the data from these new large-scale projects, it is imperative that we continue to build on the standards that have already been developed.

AUTHORS AND AFFILIATIONS
Jon Teague, The Wellcome Trust Sanger Institute; Adam Butler, The Wellcome Trust Sanger Institute; Simon Forbes, The Wellcome Trust Sanger Institute Sally Bamford, The Wellcome Trust Sanger Institute; Charlotte Cole, The Wellcome Trust Sanger Institute; David Beare, The Wellcome Trust Sanger Institute; Andrew Menzies, The Wellcome Trust Sanger Institute; Chai Yin Kok, The Wellcome Trust Sanger Institute; Rebecca Shepherd, The Wellcome Trust Sanger Institute; Nidhi Bindal, The Wellcome Trust Sanger Institute; Kenric Leung, The Wellcome Trust Sanger Institute; Mingming Jia, The Wellcome Trust Sanger Institute; David Jones, The Wellcome Trust Sanger Institute; Catherine Leroy, The Wellcome Trust Sanger Institute; John Marshall, The Wellcome Trust Sanger Institute; Keiran Raine, The Wellcome Trust Sanger Institute; Lucy Stebbings, The Wellcome Trust Sanger Institute; Peter Campbell, The Wellcome Trust Sanger Institute; Michael Stratton, The Wellcome Trust Sanger Institute; Andrew Futreal, The Wellcome Trust Sanger Institute